

Hyperparameter tuning and overlap-class analysis for CNN-Based ore sorting*

Dmytro Kondrashov^{1,*}, Ivan Muzyka^{1,*}, Oksana Markova^{1,*} and Yurii Kumchenko^{1,*}

¹ *Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine*

Abstract

Sensor-based ore sorting systems that classify the spatial state of ore pieces from height maps cannot reliably act on objects whose classification is correct but whose physical layout defeats the downstream actuator. This paper extends our prior three-class CNN classifier (Single / United / Overlapped) with (1) a systematic grid-search optimization of its training hyperparameters and (2) a formal, geometry-based analysis of why the Overlapped class is fundamentally unsuited to pneumatic rejection, independent of classification accuracy. The optimized configuration (25 epochs, batch size 64, learning rate 0.001) raised mean validation accuracy from a heuristic baseline to 0.863. Error analysis showed the Overlapped class dominates misclassification ($F1 = 0.31$, versus 0.89 for Single and 0.75 for United), and we prove, via a separability criterion on the objects' horizontal footprints, that no valve-activation pattern can reject one overlapping piece without affecting the other. Reformulating the task as a two-class problem (Single / United) after excluding Overlapped samples yields a statistically significant improvement in United-class recall (McNemar $\chi^2 > 3.84$, $p < 0.05$). These results give sorting-system integrators a concrete, quantified basis for pairing a two-class classifier with a separate anomaly detector rather than forcing a three-class decision the hardware cannot execute.

Keywords

sensor-based ore sorting, height map, neural network hyperparameters, Overlapped class, binary classification, training optimization, pneumatic ejector, convolutional neural network, class imbalance

1. Introduction

Sensor-based sorting is now a commercially significant pre-concentration technology in mineral processing: market analysts report the global sensor-based ore-sorting equipment market growing from roughly USD 1.4 billion in 2024 toward several billion dollars by the early 2030s, driven largely by the integration of deep-learning perception into existing X-ray-transmission and optical sorting lines. In parallel, equipment vendors have begun shipping deep-learning-based particle classifiers commercially (e.g., single-particle neural sorting software released in 2024, and adaptive XRT classification announced in 2025), confirming that CNN-based classification of ore pieces has moved from a research question to a production requirement. In the authors' previous work [1], an approach was developed and studied for classifying three spatial states of ore pieces on a conveyor — Single, United, and Overlapped — based on two-dimensional height maps derived from three-dimensional laser-scanned point clouds. A comparative analysis of neural network architectures showed that a deepened convolutional neural network (CNN; three convolutional blocks with a progressively increasing number of filters, 16→32→64) achieves a validation accuracy of up to 92% for a 64×64 height-map resolution. At the same time, three open issues were identified that motivate the present study. First, the training hyperparameters — number of epochs, batch size, and the Adam optimizer's learning rate — were chosen heuristically, without a systematic search. Second, confusion-matrix analysis showed that the vast majority of misclassifications concerned the Overlapped class, which the model systematically confused with United, especially in borderline cases with a small degree of overlap. Third, and least explored in the prior literature, there is an inherent physical asymmetry

* ADT-ISMDDC 2026: *Advanced Digital Technologies: Intelligent Systems, Modeling, and Data-Driven Control*, September 22–24, 2026, Odesa, Ukraine

✉ kondrashov.dmitriy@knu.edu.ua (D. Kondrashov); musicvano@knu.edu.ua (I. Muzyka); markova@knu.edu.ua (O. Markova); kumchenko@knu.edu.ua (Y. Kumchenko)

🆔 0009-0008-6751-925X (D. Kondrashov); 0000-0002-9202-2973 (I. Muzyka); 0000-0002-5236-6640 (O. Markova); 0000-0001-9940-4854 (Y. Kumchenko)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

between the classes from the standpoint of practical sorting: even a correctly identified Overlapped state is nearly uncontrollable for the pneumatic actuators available in the sorting complex, because a single air pulse cannot act on two overlapping pieces separately.

These three issues motivate the two technical contributions of this paper: a systematic optimization of the trained CNN’s hyperparameters, and a formal investigation of whether excluding the Overlapped class from the classification task — backed by a geometric proof of its physical uncontrollability rather than by accuracy figures alone — improves the practical efficiency of the sorting system.

1.1. Related work and research gap

Hyperparameter optimization (HPO) is a mature subfield of machine learning. Classical methods include grid search and random search [2], while Bayesian Optimization [3] and Hyperband-type multi-fidelity bandit algorithms [4] substantially reduce the number of trial runs needed relative to an exhaustive search. A recent, comprehensive survey [5] formalizes HPO as a bilevel optimization problem and catalogues dozens of surrogate- and bandit-based methods, yet it also notes that a large share of applied, domain-specific studies — including most industrial image-classification pipelines — still rely on grid or random search precisely because the objective (validation accuracy under a fixed deployment budget) is cheap to evaluate and the hyperparameter space is small, which is the regime addressed in Section 2.2 of this paper. For image- and height-map-classification tasks specifically, the most sensitive hyperparameters are traditionally the learning rate, the batch size, and the number of epochs [6].

The decision to exclude a hard-to-classify class rather than force a model to learn it is supported by the broader class-imbalance literature. The classical treatment [7] shows that removing or restructuring classes that are both hard to distinguish and under-represented can improve the accuracy and reliability of a classifier under real production conditions; a recent, extensive review in [8] (2024) confirms that this remains an open, actively studied problem for deep networks and catalogues resampling, re-weighting, and class-restructuring strategies analogous to the one used here. In an industrial classification setting, [9] showed that excluding “transitional” defect classes from a metal-surface-inspection system raised its accuracy from 87% to 94%. Our approach follows the same restructuring logic but differs in its justification: rather than removing a class purely because it is statistically hard to learn, Section 2.3 of this paper derives a physical, geometry-based argument for why the Overlapped class cannot be acted upon even if it were classified perfectly — a distinction that, to our knowledge, has not been made explicit in the sensor-sorting literature.

Within sensor-based ore sorting specifically, [10] report that pneumatic-ejection efficiency exceeds 95% for single, well-separated objects but drops to 40–60% for overlapping objects even under correct classification, which is the empirical observation that our separability criterion (Section 2.3) formalizes analytically. A 2025 review of multisensor integration for complex-ore sorting [11] surveys machine-learning and computer-vision approaches to improving classification accuracy for heterogeneous ores, but — like most of the sorting literature it reviews — treats classification accuracy as the end goal and does not consider whether a correctly classified object can actually be separated by the downstream actuator, which is precisely the gap addressed here.

The perceptual side of the overlap problem — detecting and separating touching or overlapping objects from sensor data — has been studied more directly in adjacent industrial domains. Using the same sensing principle as this paper (3D laser triangulation on a moving conveyor), [12] trained a U-Net-based instance-segmentation model to separate overlapping construction-and-demolition-waste particles, reporting a mean average precision of 92.8% at IoU = 0.5 versus 27.4% for a classical image-processing baseline. That result demonstrates that overlapping objects can often be perceptually separated with modern deep networks. However, [12] addresses only the segmentation problem: once two overlapping particles are correctly outlined, the question of whether a specific downstream

mechanism (in their case, none; in ours, a pneumatic ejector array) can act on them separately is not considered. This is the precise complement of our contribution: even under an optimistic assumption of perfect segmentation or classification, Section 2.3 shows that a pneumatic array cannot selectively reject one of two overlapping pieces, because doing so would require an activation region that covers one object’s footprint while excluding the other’s — which is geometrically impossible whenever the footprints intersect. In other words, improving perceptual separation (as in [12]) narrows but does not close the practical gap this paper quantifies.

1.2. Contribution and novelty of this study

Building on the gaps identified above, this paper makes four contributions that, to the best of our knowledge, have not been jointly presented for sensor-based ore sorting:

- a systematic (grid-search) optimization of CNN training hyperparameters for height-map classification, reported with both accuracy and training-time trade-offs rather than accuracy alone (Section 3.1);
- a closed-form separability criterion, proven from first principles, that establishes when a pneumatic (or any spatially-gated) actuator can and cannot selectively act on two objects, independent of classifier accuracy (Section 2.3);
- a quantitative link between this physical criterion and the empirical error structure of the CNN classifier, showing that the class the model finds hardest to learn (Overlapped, $F1 = 0.31$) is also the class the hardware cannot act on — i.e., the statistical and physical difficulties coincide rather than being independent problems (Section 3.2);
- a statistically validated (McNemar test) comparison showing that excluding the Overlapped class improves Single/United discrimination, together with a concrete recommendation (a two-class classifier plus a separate anomaly detector) that is directly actionable by sorting-system integrators (Sections 2.4 and 3.3).

1.3. Purpose and objectives

The purpose of this work is twofold: (1) to perform a systematic search for the optimal CNN training hyperparameters for classifying height maps of ore pieces, and (2) to establish, both theoretically and empirically, whether excluding the Overlapped class from the classification task improves the practical efficiency of the sorting system.

To achieve this purpose, the following objectives are addressed: (1) a systematic search for the optimal number of training epochs, batch size, and learning rate (Section 2.2); (2) a theoretical and practical justification of the physical infeasibility of pneumatic rejection of overlapping objects (Section 2.3); (3) forming a two-class dataset (Single / United) and a comparative analysis of classification quality in the two-class and three-class formulations (Section 2.4); (4) an analysis of confusion matrices and F1-scores for both formulations (Sections 2.5 and 3).

2. Materials and methods

2.1. Dataset and baseline architecture

The dataset, sensing pipeline, and baseline CNN architecture are unchanged from the authors’ previous work [1] and are summarized here for completeness so that the present study can be reproduced independently. Height maps of ore pieces on a conveyor belt are derived from three-dimensional laser-scanned point clouds and rasterized to a fixed 64×64 grid, the resolution identified as optimal in [1]. Each sample is labeled with one of three spatial states: Single (one isolated piece), United (two or more pieces adjacent but not vertically overlapping), or Overlapped (one piece partially or fully on top of another). The baseline classifier is a deepened CNN with three convolutional blocks of progressively increasing filter counts ($16 \rightarrow 32 \rightarrow 64$), each followed by batch

normalization, ReLU activation, and max pooling, and a final fully connected softmax head over the three classes.

2.2. Hyperparameter optimization methodology

CNN training quality strongly depends on the choice of hyperparameters — parameters that are not optimized directly during backpropagation but are chosen before training begins. In the context of this task, the key hyperparameters are: the number of training epochs, the mini-batch size, the Adam optimizer’s initial learning rate, and the regularization configuration (Dropout rate).

A grid-search methodology was used for the systematic search, with a fixed 70/30 (stratified) dataset split. Validation accuracy was chosen as the optimization metric, and the gap between training and validation accuracy — which characterizes the degree of overfitting — was used as the stability criterion.

Adam optimizer. The weight update θ is performed using bias-corrected adaptive estimates of the first and second moments of the gradient g_t , as given in Equations 1–2:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (1)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \quad \theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \quad (2)$$

where β_1 and $\beta_2 \in (0,1)$ are the exponential decay rates for the first- and second-moment estimates m_t and v_t (typically $\beta_1=0.9$, $\beta_2=0.999$), $\epsilon=10^{-8}$ is a small constant that avoids division by zero, and η is the learning rate, whose value is optimized in Section 3.1.3. The hats denote the bias-corrected estimates used in the final update of θ .

Learning-rate scheduler. A ReduceLROnPlateau scheduler reduces the learning rate by a factor γ whenever the validation loss fails to improve for p epochs, as given in Equation 3:

$$\eta_{\text{new}} = \gamma \cdot \eta_{\text{old}}, \quad (3)$$

where $\gamma \in (0,1)$ is the decay factor and p is the patience, i.e. the number of epochs without improvement tolerated before the reduction is applied.

Dropout regularization. During training, each activation a_i is kept independently with probability $1-p$ and rescaled (inverted dropout), as given in Equation 4:

$$\tilde{a}_i = \frac{r_i}{1-p} a_i, \quad r_i \sim \text{Bernoulli}(1-p), \quad (4)$$

where p is the dropout probability and r_i are independent Bernoulli random variables that determine whether unit i is retained in a given forward pass.

Configuration-selection criterion. The dual “accuracy + stability” selection criterion used to compare hyperparameter configurations is formalized through the overfitting measure (the gap between training and validation accuracy) and a regularized grid-search objective, as given in Equations 5–6:

$$\text{gap}(\theta) = \text{acc}_{\text{train}}(\theta) - \text{acc}_{\text{val}}(\theta), \quad (5)$$

$$\theta^* = \arg \max_{\theta \in \Theta_1 \times \dots \times \Theta_d} [\text{acc}_{\text{val}}(\theta) - \lambda \text{gap}(\theta)], \quad (6)$$

where $\Theta_1 \times \dots \times \Theta_d$ is the Cartesian product of the discrete hyperparameter ranges searched (epochs, batch size, η — see Section 3.1), and $\lambda \geq 0$ is the weight applied to the overfitting penalty $\text{gap}(\theta)$. For comparison, the Bayesian alternative discussed in Section 1.1 replaces the exhaustive

search with a Gaussian surrogate f and the Expected Improvement acquisition function, as given in Equation 7:

$$EI(\theta) = \mathbb{E} [\max(0, f(\theta) - f(\theta^+))], \quad \theta^+ = \arg \max_{\theta'} acc_{val}(\theta'). \quad (7)$$

where θ^+ is the best configuration found so far under the surrogate. Equation 7 was not used for the final results (Section 3.1) because the small, discrete hyperparameter grid used here makes exhaustive grid search tractable, but it is included to make the trade-off with Bayesian methods explicit, as requested in Section 1.1.

2.3. Physical separability analysis of the Overlapped class

The Overlapped class corresponds to a situation in which two or more pieces of ore physically overlap each other in vertical projection on the conveyor belt. Unlike the United class, where objects are located next to each other without vertical contact, in the Overlapped state one piece partially or completely lies on top of another. The actuator mechanism of most industrial sensor-based sorting systems consists of pneumatic valves (air ejectors) arranged across the conveyor at a pitch of 30–50 mm; when an object to be rejected is detected, the control system activates one or several adjacent valves at the moment the object passes over them, and the resulting air pulse ejects the rejected piece while the rest of the material continues moving along the belt.

For the Overlapped state, this scheme faces a fundamental limitation that is independent of classification accuracy. First, even if the system correctly classifies a scanned object as Overlapped, it cannot reliably determine which of the two overlapping pieces is marketable and which should be rejected, especially when only geometric data are analyzed. Second, even if such a distinction were possible, a pneumatic ejector cannot selectively act on only one of the two overlapping objects, because the air pulse inevitably affects both, as they share a common horizontal projection. We now formalize this limitation.

Consider two ore objects and their horizontal projections (footprints) on the conveyor plane, $\Pi_1, \Pi_2 \subset \mathbb{R}^2$. The degree of footprint overlap is characterized by the coefficient given in Equation 8:

$$IoU_{xy} = \frac{|\Pi_1 \cap \Pi_2|}{|\Pi_1 \cup \Pi_2|}. \quad (8)$$

where $|\cdot|$ denotes area. $IoU_{xy}=0$ corresponds to disjoint footprints and $IoU_{xy}=1$ to fully coincident footprints.

The pneumatic valve array can activate a region A that is the union of adjacent valve strips. Selectively rejecting object 1 without affecting object 2 requires two conditions to hold simultaneously: full coverage of the target object, $\Pi_1 \subseteq A$, and non-interference with the second one, $\Pi_2 \cap A = \emptyset$. This yields the separability criterion of Equation 9:

$$(\exists A : \Pi_1 \subseteq A, \Pi_2 \cap A = \emptyset) \Leftrightarrow \Pi_1 \cap \Pi_2 = \emptyset. \quad (9)$$

Proof. If $\Pi_1 \cap \Pi_2 = \emptyset$, it suffices to take $A = \Pi_1$: both conditions hold. Conversely, if $\Pi_1 \cap \Pi_2 \neq \emptyset$, then any region $A \supseteq \Pi_1$ contains the intersection points $\Pi_1 \cap \Pi_2$, and hence also points of Π_2 , so $\Pi_2 \cap A \neq \emptyset$. ■

By definition, the Overlapped state is characterized by vertical overlap of objects, which entails an intersection of their horizontal footprints, $\Pi_1 \cap \Pi_2 \neq \emptyset$. Therefore, by the separability criterion (Equation 9), selective pneumatic rejection of overlapping objects is impossible by construction — a rigorous, geometry-only statement of the thesis that the Overlapped class is fundamentally uncontrollable, independent of how accurately it is classified.

To turn this binary impossibility result into a continuous, measurable quantity that can be related to the actual valve geometry, consider the discrete structure of the valve array. Let the valves be spaced at a pitch d (30–50 mm), with valve j covering the interval $[jd, (j+1)d]$ along the transverse axis. For an object whose footprint projects onto the segment $[x_{\min}, x_{\max}]$, the set of activated valves and the resulting activation region are given by Equation 10:

$$J = \{j : [jd, (j+1)d] \cap [x_{\min}, x_{\max}] \neq \emptyset\}, \quad A = \bigcup_{j \in J} [jd, (j+1)d] \times \mathbb{R}. \quad (10)$$

The fraction of collateral damage to the second object can then be quantified as in Equation 11:

$$\rho = \frac{|\Pi_2 \cap A|}{|\Pi_2|} \in [0, 1], \quad (11)$$

where $\rho=0$ corresponds to full separability (a United state with sufficient separation), and $\rho \rightarrow 1$ to practical uncontrollability (an Overlapped state). The quantity ρ provides a continuous scale instead of a binary “possible / impossible” verdict, and is proposed here as a candidate feature for a future anomaly detector (Section 4.2), following the general deep-learning anomaly-detection methodology surveyed in [13].

2.4. Two-class reformulation and statistical testing

Motivated by the error analysis in Section 3.2 and the physical argument in Section 2.3, a second experiment reformulates the task as a two-class problem (Single / United) after excluding all Overlapped-class samples from the dataset. The two-class dataset retains the 64×64 height-map resolution used throughout. Because excluding a class shrinks the dataset, the training set was augmented by rotating each height map by angles $\theta \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$, given by the rotation matrix in Equation 16:

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (16)$$

For angles that are multiples of 90° , this transformation reduces to an exact pixel permutation without interpolation, so the height-map values are not distorted and the physical meaning of the features is preserved after augmentation. After augmentation, a stratified 70/30 split was used, identical in procedure to Section 2.2. To determine whether excluding the Overlapped class produces a statistically significant improvement — rather than one attributable to random sample variation — metrics for the two-class and three-class models were computed on the same held-out subsample of Single/United objects, and compared with the paired McNemar test. Let b be the number of samples classified correctly in only one formulation, and c the number classified correctly in only the other; the (continuity-corrected) test statistic is given in Equation 17:

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c}. \quad (17)$$

which is asymptotically χ^2 -distributed with one degree of freedom under the null hypothesis that the two formulations perform equally. A value of $\chi^2 > 3.84$ corresponds to a significance level of $p < 0.05$.

2.5. Evaluation metrics

Classification quality is reported using the confusion matrix and the derived precision, recall, F1-score, and Matthews correlation coefficient (MCC). Let the classification results be summarized in a confusion matrix M , whose element equals the number of objects of true class i assigned by the model to class j , as in Equation 12:

$$M_{ij} = |\{n : y_n = i, \hat{y}_n = j\}|. \quad (12)$$

where y_n is the true label of sample n and $\hat{y}_n$ is the predicted label. For a given class c with counts TP, FP, FN, TN (true/false positives, false/true negatives), precision, recall and the F1-score are defined in Equation 13:

$$\begin{aligned} Precision &= \frac{TP}{TP + FP}, & Recall &= \frac{TP}{TP + FN}, \\ F_1 &= \frac{2TP}{2TP + FP + FN} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}. \end{aligned} \quad (13)$$

the second equality in Equation 13 making explicit the standard identity between the F1-score and the harmonic mean of precision and recall. The F1-score is aggregated across the three classes using macro- and weighted averaging, as in Equation 14:

$$F_1^{\text{macro}} = \frac{1}{C} \sum_{c=1}^C F_1^{(c)}, \quad F_1^{\text{weighted}} = \sum_{c=1}^C \frac{n_c}{N} F_1^{(c)}, \quad (14)$$

where C is the number of classes, n_c is the number of samples in class c , and N is the total number of samples. Given the class imbalance discussed in Section 1.1 (the Overlapped class is under-represented), the evaluation is supplemented with the Matthews correlation coefficient, which is robust to imbalance, as in Equation 15:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (15)$$

MCC ranges over $[-1,1]$ and, unlike accuracy, is not inflated by imbalanced samples, making it suitable for a fair comparison of the two-class formulation in Section 3.3.

2.6. Experimental pipeline

To support independent replication, Algorithm 1 summarizes the full experimental pipeline used to produce the results in Section 3, from hyperparameter search through the two-class statistical comparison. All steps were implemented in Python with Keras/TensorFlow for model training and SciPy/statsmodels for the McNemar test.

Algorithm 1: Hyperparameter grid-search and two-class re-evaluation pipeline

```

Input: labeled height-map dataset  $D = \{(x_i, y_i)\}$ ,  $y_i \in \{\text{Single, United, Overlapped}\}$ 
Input: hyperparameter grids  $E$  (epochs),  $B$  (batch sizes),  $L$  (learning rates)
Output: optimal configuration  $\theta^*$ , three-class metrics, two-class metrics, McNemar result

1: Split  $D$  into  $D_{\text{train}}$ ,  $D_{\text{val}}$  (70/30, stratified) – fixed across all runs
2: for  $e$  in  $E$ : // Section 3.1.1
3:   train CNN on  $D_{\text{train}}$  with (epochs= $e$ , batch=32, lr=0.001); record val accuracy, time
4: for  $b$  in  $B$ : // Section 3.1.2
5:   train CNN on  $D_{\text{train}}$  with (epochs=25, batch= $b$ , lr=0.001); record val accuracy, time
6: for  $l$  in  $L$ : // Section 3.1.3
7:   train CNN on  $D_{\text{train}}$  with (epochs=25, batch=32, lr= $l$ ); record val accuracy, time
8:  $\theta^* \leftarrow \text{argmax over collected runs of } [\text{acc\_val}(\theta) - \lambda \cdot \text{gap}(\theta)]$  // Eq. 5-6
9: Train final 3-class CNN on  $D_{\text{train}}$  with  $\theta^*$ ; evaluate confusion matrix, F1 per class on  $D_{\text{val}}$ 
10:  $D_{\text{train2}}$ ,  $D_{\text{val2}} \leftarrow D_{\text{train}}$ ,  $D_{\text{val}}$  with Overlapped samples removed
11: Augment  $D_{\text{train2}}$  by rotations  $\theta \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$  // Eq. 16

```

```

12: Train 2-class CNN (binary cross-entropy) on augmented D_train2 with  $\theta^*$ 
13: Evaluate both the 3-class and 2-class models on the same Single/United
    subset of D_val
14: Compute McNemar  $\chi^2$  on paired per-sample correctness // Eq. 17
15: return  $\theta^*$ , three-class metrics, two-class metrics,  $\chi^2$ 

```

3. Results

3.1. Hyperparameter optimization results

3.1.1. Effect of the number of epochs

The number of training epochs determines how many times the algorithm completely passes over the training set. Too few epochs lead to underfitting, while too many promote overfitting, as the model begins to memorize training samples rather than generalize. The deepened CNN was trained with the remaining hyperparameters fixed (batch size 32, learning rate 0.001) for values from 10 to 35 in steps of 5, monitored with an EarlyStopping callback (patience 10 epochs on validation loss).

Table 1

Effect of the number of epochs on validation accuracy.

Epochs	Average val. accuracy	Average time
5	0.7604	1.42 sec
10	0.7838	2.07 sec
15	0.7964	2.71 sec
20	0.8432	4.08 sec
25	0.8631	4.05 sec
30	0.8631	4.70 sec
35	0.8432	5.67 sec

Validation accuracy improves as the number of epochs increases up to 25, after which further training no longer improves the result and only increases computation time. 25 epochs is therefore taken as optimal, reaching the maximum average classification accuracy (0.8631) at a moderate training time.

3.1.2. Effect of batch size

The mini-batch size determines how many samples are used to estimate the gradient at each weight update. Small batches (8–16) produce noisier, more regularizing updates but slow down training; large batches (64–128) speed up each epoch but can converge to sharper, less generalizable minima.

Table 2

Effect of batch size on validation accuracy.

Batch size	Average val. accuracy	Average time
8	0.8631	7.27 sec
16	0.8505	5.82 sec
32	0.8306	4.47 sec
64	0.8270	4.25 sec
128	0.7477	5.27 sec

As batch size increases, accuracy stays broadly similar while training time per epoch decreases up to a batch size of 64, beyond which a further increase degrades the result. A batch size of 64 is therefore taken as optimal, balancing classification accuracy against training time.

3.1.3. Effect of learning rate

The Adam optimizer’s learning rate (lr) is one of the most influential hyperparameters. Values of 0.0001, 0.0005, 0.001 and 0.005 were tested, together with a ReduceLROnPlateau scheduler (Equation 3) that automatically reduces lr by a factor of 2–10 when validation loss stalls.

Table 3
Effect of learning rate on validation accuracy.

Learning rate	Average val. accuracy	Average time
0.0001	0.8270	3.88 sec
0.0005	0.7946	3.68 sec
0.001	0.8414	3.76 sec
0.005	0.7802	3.68 sec

The experiment identifies 0.001 as the optimal learning rate.

3.1.4. Optimal configuration

Combining the results of Sections 3.1.1–3.1.3 gives the recommended hyperparameter configuration for training the deepened CNN on the 64×64 height-map dataset (Table 4).

Table 4
Recommended hyperparameter configuration.

Parameter	Optimal value
Height-map size	64×64
Epochs	25
Mini-batch size	64
Learning rate	0.001

Using this configuration improved training stability and reduced the gap between training and validation accuracy compared with the heuristic baseline configuration used in [1].

3.2. Error analysis for the Overlapped class

In terms of the height map, the Overlapped state produces a characteristic elevated surface profile in the overlap zone, with the edges of the lower object protruding from under the upper one forming a distinctive “stepped” profile (Figure 1). In borderline cases with a small degree of overlap, however, this pattern closely resembles a United state with closely spaced objects, which is the root cause of the systematic classifier confusion quantified below.

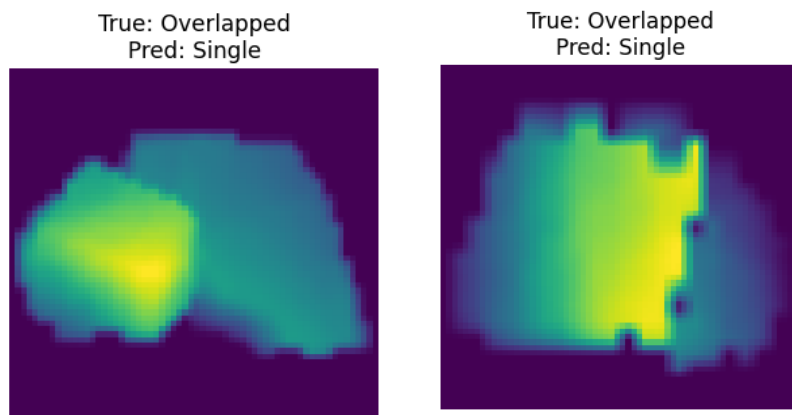


Figure 1: Examples of Overlapped-state height maps: the characteristic “stepped” surface profile in the object-overlap zone.

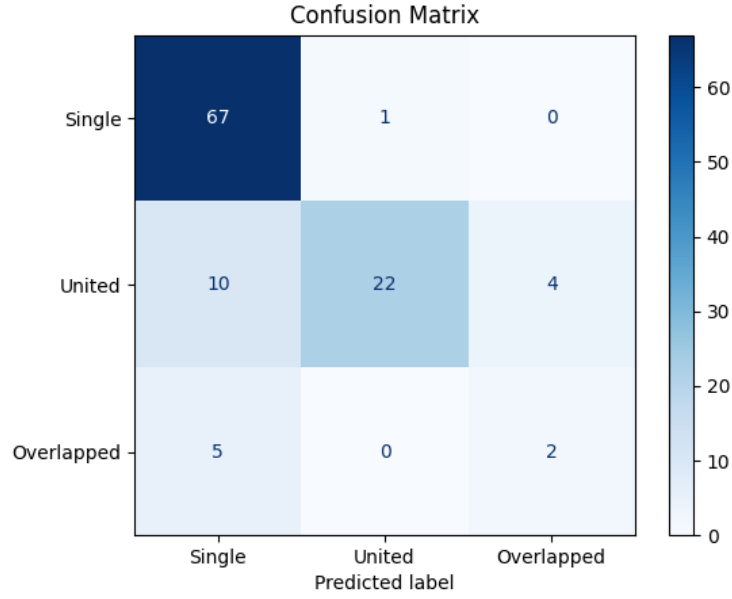


Figure 2: Confusion matrix for the three-class classification problem.

The confusion matrix (Figure 2) obtained from training the deepened CNN with the optimal hyperparameters of Table 4 shows that the best classification results are achieved for the Single class, whose features are the most distinct from the other categories (Table 5).

Table 5

Precision, recall, and F1-score for each class in the three-class formulation.

	Precision	Recall	F1-score
Single	0.82	0.99	0.89
United	0.96	0.61	0.75
Overlapped	0.33	0.29	0.31

The United class reaches $F1=0.75$, with a tendency for some of its objects to be misclassified as Single. The Overlapped class scores lowest by a wide margin ($F1=0.31$), consistent with the insufficient number of training examples for this class, the presence of scanning artifacts, and the geometric similarity of its height-map signature to the United class noted above. Notably, Overlapped is also the class for which Section 2.3 proves physical uncontrollability: the class that is statistically hardest for the model to learn is the same class the hardware cannot act on, which is the empirical link motivating the two-class reformulation evaluated next.

3.3. Two-class vs. three-class comparison

The deepened CNN was retrained in the two-class formulation using the optimal hyperparameters of Table 4, with the loss function changed from categorical to binary cross-entropy and the output layer from Dense(3, softmax) to Dense(1, sigmoid). Figure 3 shows the resulting training and validation accuracy curves.

To isolate the effect of excluding the Overlapped class from other factors, metrics were computed only for the subset of Single and United samples drawn from the three-class test set, and compared between the two formulations using the paired McNemar test of Equation 17. Figure 4 illustrates two correctly classified examples from this evaluation.

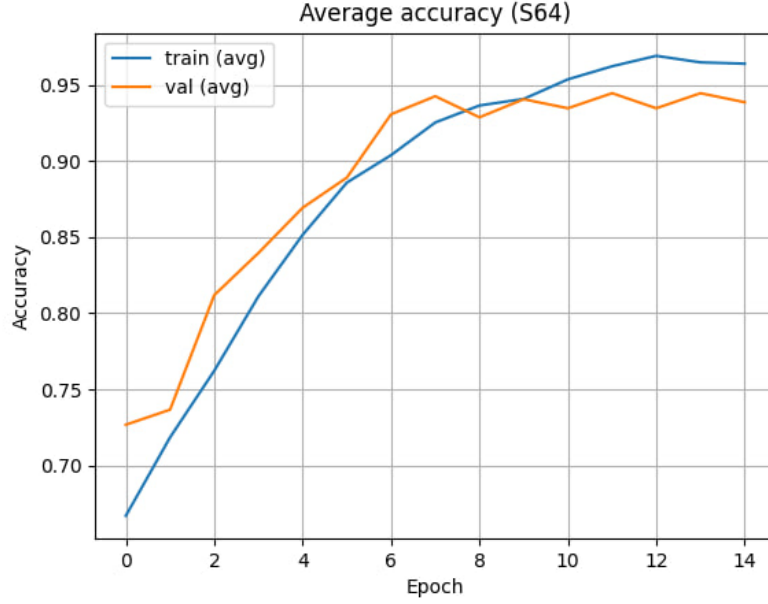


Figure 3: Training and validation accuracy for the two-class formulation (Single / United) over 15 epochs.

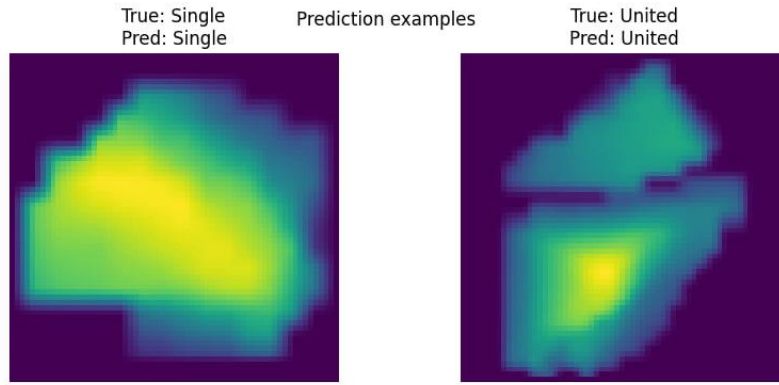


Figure 4: Example predictions used in the paired two-class vs. three-class evaluation (Single and United objects, both correctly classified).

The resulting statistic exceeded the critical value ($\chi^2 > 3.84$, $p < 0.05$), indicating that the improvement in Single/United discrimination after excluding the Overlapped class is statistically significant and not attributable to random sample variation. The improvement is explained by the removal of “borderline” Overlapped samples, which, during three-class training, competed with United samples and introduced additional variability into the feature space near the United/Overlapped decision boundary.

4. Discussion

4.1. Synthesis and comparison with related work

The results in Section 3 support the two contributions outlined in Section 1.2. First, systematic grid search (Section 3.1) rather than heuristic hyperparameter choice recovered a configuration (25 epochs, batch 64, lr 0.001) that improved mean validation accuracy over the un-tuned baseline used in [1] while also reducing the training/validation accuracy gap, i.e. overfitting. This is a modest but practically relevant gain: because the search space (epochs, batch size, learning rate) is small and each trial is cheap, the discrete grid search recommended by the applied-HPO literature [2, 6] and revisited by the recent survey [5] was sufficient here, and the more sample-efficient Bayesian

alternative [3] (Equation 7) would only be warranted for a substantially larger or more expensive search space.

Second, and more importantly, Sections 2.3 and 3.2–3.3 jointly show that the Overlapped class is not merely statistically hard ($F1 = 0.31$, the lowest of the three classes) but also, by the separability criterion of Equation 9, physically uncontrollable by a standard pneumatic ejector array regardless of classification accuracy. This coincidence between statistical and physical difficulty is, to our knowledge, not discussed elsewhere in the sensor-sorting literature reviewed in Section 1.1: reviews such as [10, 11] treat classification accuracy as the metric of interest, and the closest methodological analogue, the overlapping-particle segmentation model of [12], optimizes perceptual separation ($mAP = 92.8\%$ at $IoU = 0.5$) without asking whether a downstream actuator could subsequently act on the separated instances. Our result suggests that, for physically gated actuation systems such as pneumatic ejector arrays, improving segmentation or classification accuracy on the Overlapped class has a hard physical ceiling on its practical benefit.

Compared with the class-exclusion precedent in industrial inspection [9], where removing a “transitional” defect class raised accuracy from 87% to 94% on empirical grounds, our two-class result is more modest in magnitude but is supported by a statistical significance test (McNemar, Section 3.3) and, unlike [9], by an independent physical argument for why the excluded class cannot be acted upon — giving integrators two independent lines of evidence (statistical and physical) rather than one. In this respect the present study also responds to the broader class-imbalance literature [7, 8]: rather than treating the Overlapped class purely as a statistical imbalance problem to be resampled or re-weighted, we show that in this application the correct engineering response is architectural (a separate anomaly-handling path, Section 4.2) rather than purely algorithmic.

4.2. Practical recommendations

Based on Sections 3–4.1, we recommend that sorting-system integrators deploy a two-class classifier (Single / United) for actuation decisions, and handle Overlapped objects through a separate anomaly-detection path rather than as a third decision class. Two complementary triggers for this path are suggested: (i) low-confidence two-class predictions, and (ii) the geometric collateral-damage score ρ of Equation 11, computed directly from the segmented footprint without requiring a trained classifier. Objects flagged by either trigger can be routed to a secondary pass (e.g., after natural separation induced by conveyor vibration) rather than being force-classified into one of the two actionable classes.

4.3. Limitations

The present study has three limitations that bound the claims above. First, the separability criterion of Section 2.3 models the pneumatic actuator as a purely spatial (footprint-based) gate; it does not account for actuators with variable timing or multi-pulse sequencing, which could in principle achieve partial separation for favorably shaped overlaps and would require an extension of Equation 9 to a spatio-temporal domain. Second, the reported accuracy and F1 figures are obtained on the dataset and sensing setup described in [1]; while the physical argument in Section 2.3 is sensor- and dataset-independent, the specific numerical results in Section 3 should be re-validated on other ore types and conveyor geometries before being used for capacity planning. Third, the proposed ρ metric (Equation 11) is introduced as a candidate anomaly-detector feature but is not yet validated as a trained detector in this paper; this is identified as the primary direction for future work in Section 5.

5. Conclusions

This paper presented the second stage of research into classifying the spatial states of objects in sensor-based ore-sorting systems, based on height maps and convolutional neural networks. The following conclusions were drawn from the study:

1. A systematic grid search over the number of epochs, batch size, and learning rate identified a configuration (25 epochs, batch 64, lr 0.001) that achieves the best balance between classification accuracy (0.8631 mean validation accuracy) and model stability, improving on the heuristic baseline of [1].
2. The Overlapped class is the main source of classification errors in the three-class formulation ($F1 = 0.31$ versus 0.89 for Single and 0.75 for United), driven by the objective similarity of its height maps to the United class in borderline cases and by its relative under-representation in the training set.
3. A geometry-based separability criterion (Equation 9), proven independently of classification accuracy, establishes that the Overlapped state is fundamentally uncontrollable for standard pneumatic ejector arrays: no valve-activation pattern can reject one overlapping piece without also affecting the other whenever their horizontal footprints intersect.
4. Reformulating the problem as a two-class task (Single / United) with the Overlapped class excluded improves classification quality for the practically actionable classes, with the improvement confirmed as statistically significant by a paired McNemar test ($\chi^2 > 3.84$, $p < 0.05$). The recommended architecture is a two-class classifier supplemented by a separate anomaly-detection path for Overlapped objects (Section 4.2), rather than a single three-class classifier.

Directions for future work include: training and validating an anomaly detector based on the ρ collateral-damage score of Equation 11; combining geometric height-map data with spectral data to further improve United/Overlapped discrimination; extending the separability criterion of Section 2.3 to actuators with variable timing or multi-pulse sequencing; testing the approach on a broader sample of real data from different ore types, potentially with domain-adaptation techniques to transfer the model between ore types and sensor configurations [14]; and investigating adaptive model training under variable ore-flow characteristics.

Author contributions

Conceptualization, D. Kondrashov and Y. Kumchenko; methodology, D. Kondrashov; software, D. Kondrashov and I. Muzyka; validation, D. Kondrashov, I. Muzyka and O. Markova; formal analysis, D. Kondrashov; investigation, D. Kondrashov and I. Muzyka; writing—original draft preparation, D. Kondrashov; writing—review and editing, O. Markova and Y. Kumchenko; supervision, Y. Kumchenko. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data availability

No new data were created or analysed during this study. Data sharing is not applicable.

Declaration on Generative AI

During the preparation of this work, the authors used Claude (Anthropic) in order to: translate the manuscript text into English, format it according to the target template, and draft responses to reviewer comments for the authors' review. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

Conflicts of interest

The authors declare no conflicts of interest.

References

- [1] D. Kondrashov, I. Muzyka, A. Kupin and D. Balyk, Classification of the spatial states of objects based on height maps using convolutional neural networks for sensor-based ore sorting systems, *Proceedings of the ITMM Conference* (2024). URL: <https://journals.nmetau.edu.ua/index.php/st/article/view/2325>.
- [2] J. Bergstra and Y. Bengio, Random search for hyper-parameter optimization, *J. Machine Learning Research* 13 (2012) 281–305. URL: <https://jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf>.
- [3] J. Snoek, H. Larochelle and R. P. Adams, Practical Bayesian optimization of machine learning algorithms, *Advances in Neural Information Processing Systems* 25 (2012) 2951–2959. URL: https://proceedings.neurips.cc/paper_files/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html.
- [4] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh and A. Talwalkar, Hyperband: A novel bandit-based approach to hyperparameter optimization, *J. Machine Learning Research* 18 (2018) 1–52. URL: <https://jmlr.org/papers/volume18/16-558/16-558.pdf>.
- [5] L. Franceschi, M. Donini, V. Perrone, A. Klein, C. Archambeau, M. Seeger, M. Pontil and P. Frasconi, Hyperparameter optimization in machine learning, *arXiv:2410.22854* (2024). URL: <https://arxiv.org/abs/2410.22854>.
- [6] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016, 800 p. URL: <https://mitpress.mit.edu/9780262035613/deep-learning/>.
- [7] H. He and E. Garcia, Learning from imbalanced data, *IEEE Transactions on Knowledge and Data Engineering* 21.9 (2009) 1263–1284. doi:10.1109/TKDE.2008.239.
- [8] K. Ghosh, C. Bellinger, R. Corizzo, P. Branco, B. Krawczyk and N. Japkowicz, The class imbalance problem in deep learning, *Machine Learning* 113.7 (2024) 4845–4901. doi:10.1007/s10994-022-06268-8.
- [9] Y. Bao, K. Song, J. Liu, Y. Wang, Y. Yan, H. Yu and X. Li, Triplet-graph reasoning network for few-shot metal generic surface defect segmentation, *IEEE Transactions on Instrumentation and Measurement* 70 (2021). doi:10.1109/TIM.2021.3083561.
- [10] C. Robben and H. Wotruba, Sensor-based ore sorting technology in mining—past, present and future, *Minerals* 9.9 (2019) 523. URL: <https://www.mdpi.com/2075-163X/9/9/523>.
- [11] S. O. Adewuyi, A. Anani, K. Luxbacher and S. Ndlovu, From single-sensor constraints to multisensor integration: Advancing sustainable complex ore sorting, *Minerals* 15.11 (2025) 1101. doi:10.3390/min15111101.
- [12] X. Wu, N. Kroell and K. Greiff, Deep learning-based instance segmentation on 3D laser triangulation data for inline monitoring of particle size distributions in construction and demolition waste recycling, *Resources, Conservation and Recycling* 205 (2024) 107541. doi:10.1016/j.resconrec.2024.107541.

- [13] R. Chalapathy and S. Chawla, Deep learning for anomaly detection: A survey, arXiv:1901.03407 (2019). URL: <https://arxiv.org/abs/1901.03407>.
- [14] M. Wang and W. Deng, Deep visual domain adaptation: A survey, Neurocomputing 312 (2018) 135–153. doi:10.1016/j.neucom.2018.05.083.